



A Study on Non-negative Matrix Factorization under Probability Constraints

| | |
|----------|---|
| 著者 | 伊藤 寛祥 |
| 発行年 | 2020 |
| その他のタイトル | 確率値制約のもとでの非負値行列分解に関する研究 |
| 学位授与大学 | 筑波大学 (University of Tsukuba) |
| 学位授与年度 | 2019 |
| 報告番号 | 12102甲第9403号 |
| URL | http://hdl.handle.net/2241/00160952 |

| | |
|-----------|---|
| 氏 名 | 伊藤 寛祥 |
| 学 位 の 種 類 | 博 士 (工 学) |
| 学 位 記 番 号 | 博 甲 第 9 4 0 3 号 |
| 学位授与年月日 | 令和2年3月25日 |
| 学位授与の要件 | 学位規則第4条第1項該当 |
| 審 査 研 究 科 | システム情報工学研究科 |
| 学位論文題目 | A Study on Non-negative Matrix Factorization under Probability Constraints (確率値制約のもとでの非負値行列分解に関する研究) |
| 主 査 | 筑波大学 教授 博士(工学) 天笠 俊之 |
| 副 査 | 筑波大学 教授 理学博士 北川 博之 |
| 副 査 | 筑波大学 教授 博士(工学) 山本 幹雄 |
| 副 査 | 筑波大学 教授 博士(工学) 佐久間 淳 |
| 副 査 | 筑波大学 准教授 博士(情報学) 加藤 誠 (図書館情報メディア研究科) |

論 文 の 要 旨

行列分解は、行列をより低ランクの行列の積として表現あるいは近似する手法であり、これまでテキストやグラフを含む幅広い応用分野で利用されてきた。とりわけ、非負値行列分解 (non-negative matrix factorization; NMF) は、入力行列に対して各成分が非負という制約の下で行列分解を行なう手法であり、単純でありながら高い性能を示し、さらに解釈性に優れるといった性質を持つことから、近年さまざまな応用で用いられている。ところで、行列を用いた分析においては行列の各成分を確率値として扱いたい場合が少なくないが、このような行列に対して NMF を適用した場合、その結果得られる行列は必ずしも確率の制約を満たしていないことが問題となっていた。

そのため本研究では、行列の各成分が確率値となっている行列 (確率行列) に対して、確率値としての制約を満たしながら、非負値行列分解を行なう新たな手法 Probabilistic Matrix Factorization (PMF) を提案している。本研究による主な貢献は以下の通りとなっている: 1) PMF の形式的な定義を与えている、2) 交互最適化に基づく最適化アルゴリズムを示すとともに、その計算量が既存の NMF と等しいことを示している、3) PMF をより一般的な四つのパターンの確率行列に拡張した手法を提案している、4) PMF の特定のケースが既存手法である pLSA あるいは LDA と等価であることを明らかにしている、5) 複数のテキストデータセットを用いた実験を行ない、クラスタリングにおいて PMF が従来手法を上回る性能を達成していることを示している。

上記に加えて、PMF の有効性を示すことを目的として、複数の最適化問題を同時に解くマルチタスク学習によるグラフ分析にも取り組んでいる。実世界には、各頂点が複数のテキスト属性で特徴づけられているグラフが多数存在している。また、関連性の強い頂点集合 (コミュニティ) を検出する、

あるいはコミュニティにおいて特徴的に出現するテキストを検出することはさまざまな応用において有用である。このとき特定のコミュニティには特徴的に出現するテキストが存在することが多く、また特定のテキストによって特徴づけられる頂点はそれに対応するコミュニティと関連性が強い。このことから両者を同時に検出することで、互いの精度を向上することができる。このアイデアに基づいて、本研究では頂点のコミュニティ、コミュニティにおけるテキスト特徴および両者の関係性を複数の PMF として定式化し、それらを同時に最適化するアルゴリズムを提案している。実データを利用した実験により、提案手法が既存手法を上回る精度を達成していることを示している。

審 査 の 要 旨

【批評】

行列分解、とりわけ非負値行列分解 (NMF) は、テキストやグラフなど、幅広い分野において用いられている手法である。特に記述対象が確率的に表現される応用においては、確率値を成分とする確率行列は頻繁に用いられる。しかしながら、従来の NMF を確率行列に適用した場合、結果として得られる行列は必ずしも確率行列とはならず、既存研究ではこの問題に対してアドホックな対応をしてきた。

この問題に対して本研究では、確率の制約を満たしながら行列分解を行なうための新たな手法 PMF (Probabilistic Matrix Factorization) を提案している。そのために PMF の定式化および最適化アルゴリズムを提案するとともに、PMF の一般化や PMF の既存手法との関係性についても明らかにしている。さらに PMF の有効性を示すことを目的として、複数の属性を持つグラフデータの分析にも取り組み、ノードのコミュニティ、コミュニティのテキスト特徴および両者の関連性を高い精度で検出可能なことを実験的に示している。

このように本研究では、従来の NMF を発展させ、確率行列を確率の制約の下で直接確率行列の積に分解することができる手法を提案し、さらにその有効性をテキストデータ、属性付きグラフデータの分析を通して実験的に検証している点において、情報工学上の理論的、技術的貢献が認められる。今後は提案手法の有効性を異なる応用分野のデータで検証することが期待される。

【最終試験の結果】

令和 2 年 2 月 18 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。